Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 76 (2015) 308 – 315
Procedia Computer Science

2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

Implementation of GLCM Features in Thermal Imaging for Human Affective State Detection

Abd Latif,M.H[a], Md.Yusof,H[*], Sidek,S.N , Rusli, N

[a]*Department of Mechatronics Engineering, International Islamic University Malaysia,Kuala Lumpur-53100, Malaysia*

Abstract

Human Robot Interaction (HRI) is a multidisciplinary field which involves developing, perceiving and assessing robotic systems. To ensure the effectiveness in communication, the understanding of emotions and intentions is essential. In the development of an emotionally intelligent robot, the issues on how to perceive human affective states and how to manifest the robot's emotion should be addressed. Recently, thermal imaging has exhibits potential solution for non-invasive recording of Autonomic Nervous System (ANS). The ANS works by measuring the spontaneous thermal radiation radiated from human body. In this paper, we present an efficient method for thermal image feature extractions using the Gray Level Co-occurrence Matrix (GLCM) technique. This work attempts to investigate the suitability and sensitivity of the thermal imaging technique for affect detection by analysing the heat pattern on the facial skin. Four region of interests (ROIs), Supraorbital, Periorbital, Nasal and Mouth are looked into where the second order statistical features (Contrast, Correlation, Energy, and Homogeneity) are extracted and used to predict the emotional states.The findings of this study indicates that thermal imaging as an alternative, contactless and non-invasive method for appraising human emotional states.

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

Keywords: Affect Detection; Human Robot Interaction (HRI); Thermal Image ;Feature Extraction; Gray Level Co-occurrence Matrix (GLCM)

1. Introduction

Over the years, Human Robot Interaction (HRI) has become an emerging field of research which attracted many researches from diversify domains. The growing number of researchers involving in this particular field has

* Corresponding author. Tel.: (+603) 6196 5712
E-mail address: myhazlina@iium.edu.my

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)
doi:10.1016/j.procs.2015.12.298

demonstrated the importance of HRI. The domain of HRI started through a one-way communication using a hand controller between operator and robot where the robot was viewed as the extension of human body (Sheridan, 1992). Advancement in sensing technique, control paradigms, artificial intelligence, and computation power has opened a new dimension for HRI to exploit new possibilities.

Human behavior can be best expressed through emotional reactions and these reactions compose of wide variety of external stimuli; pleasant situation, psychological pressure etc. This emotional reaction can be manifested in countless ways; facial expressions, speech, body gestures and physiological signals (EEG, heart rate, blood flow, temperature etc.). While a lot of data and findings had been documented in the area of affect detection using Autonomic Nervous System (ANS) parameters, most of these experimentations uses invasive approaches where direct contact between the subject and the sensor is required. This would be intimidating to the subject and restricts their mobility, which somehow would contribute some uncertain information due to stress experienced by the subjects. The maturity and evolution of sensing technique has made contact-less and non-invasive ANS responses monitoring possible through the application of thermal imaging which has shown potential in characterizing the subdivision of ANS. Merla et al. has proposed the facial skin temperature exploitability for affective states recognition using thermal imaging (Merla & Romani, 2007). Merla et al. (2007) stated that "Measuring subtle thermal effect due to emotional arousal may provide useful information about the sympathetic activity". Certain areas of the facial skin have a direct correlation with affective states as being documented in several findings.

A robotic system that has the capability of recognizing emotional states and synthesizing proper response would be beneficial for Human Robot Interaction (HRI). Hence, empowering computers and robots to understand human emotions would make human robot interaction more meaningful. The potential applications of robots that could detect a person's affective states and respond accordingly to such perception are diverse (personal home aids, entertainment robot, rescue and surveillance robots etc.).

In this paper, we present our preliminary findings in affective states detection through the implementation of GLCM features for thermal images. The remaining part of this paper is organized as follows: Section 2 describes the Thermal Imaging Affect Detection. Experimental Design is presented in Section 3. Section 4 describes some processes require for implementation of thermal image for affect evaluation. Lastly, in Section 5 some discussion and finding are discussed.

2. Thermal Imaging Affect Detection

To date, observing, accessing and classifying of emotional states has been done through the mean of monitoring the autonomic nervous system (ANS) which has been proven effective yet this traditional approach requires direct contact between the subject and the sensor which potentially biasing the estimation of the state as the compliant involvement of subject is required (Merla, 2014). Numerous known modalities are available for accessing the ANS parameters; heartbeat, breathing modulation, ECG, skin conductance and others. The ANS parameters provides an effective tool for observing and classifying emotional responses as documented in countless publications. According to Merla et al., thermal imaging has manifest potential solution for non-invasive recording of ANS. It works by measuring the spontaneous thermal radiation of the body in non-invasive and contact-less manner. Based on the previous research findings in the domain of ANS, thermal imaging has demonstrated its ability to characterize the subdivision of ANS (Murthy and Pavlidis, 2006; Garbey et al., 2007; Merla and Romani, 2007; Pavlidis et al., 2007; Shastri et al., 2009; Merla, 2013; Engert et al., 2014).

Most of the research done in affective state detection using thermal imaging have chosen the face area for psychological signal measurement due to its direct involvement in social interaction and communication. Besides, the skin layer in face area is thinner and well exposes which make it more feasible for psychological signal recording. Furthermore, the validity and reliability of this method has been proven by comparing the result side by side with golden standard methods such as ECG, galvanic skin response (GSR) etc. Thermal imaging and GSR has elucidated the similar detection power as reported by several studies (Coli et al., 2007; Shastri et al., 2009; Pavlidis et al., 2012; Di Giacinto et al., 2014; Engert et al., 2014).

There are two types of features extracted from the thermal images; imaging features and temperature features (Liu & Wang, 2011). Among the researches that uses extracted imaging features from the imaging feature: Yasunari Yoshitomi et al. (2010) where the extracted features are transformed by using a two-dimensional discrete cosine transformation (2D-DCT) to transform the gray scale values in the facial area into frequency components, and used these extracted features in the expression recognition systems. Benjam´ın Hern´andez, Gustavo Olague et al. (2007) had chosen Gray Level Co-occurrence Matrix (GLCM) for region descriptors computation of the infrared images based on second order statistics features to distinguish the expressions of surprise, happiness and anger. From the temperature features; Masood Mehmood Khan et al. (2009) have exploited the variances in the thermal intensity values recorded at thermally significant locations on human faces as the features to distinguish pretended and evoked emotional expressions. Brain R. Nhan and Tom Chau (2010) have extracted the time, frequency and time-frequency features derived thermal infrared data to classify the natural responses of subject- levels of arousal and valence induced by the International Affective Picture System (IAPS) and others.

3. Experimental Design

The proposed framework of our method is shown in Fig.1 which involves three main blocks; Image Processing, Feature Extraction and Test and Validation. In the first block (Image Processing), several steps involved named image acquisition, pre-processing (Contrast Limited Adaptive Histogram Equalization (CLAHE)) and ROIs generation. The process then followed by second block (Feature Extraction) which we used the Gray Level Co-occurrence Matrix (GLCM) approach for second order statistical features extraction. Lastly, the third block (Test & Validation) involved classification and k-fold cross validation.

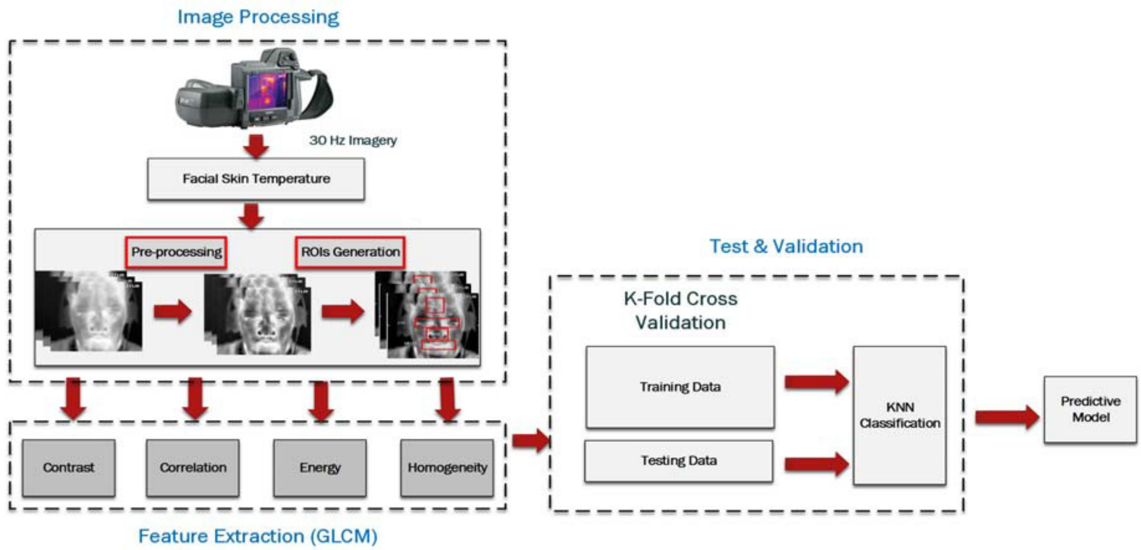

Figure 1: Proposed Framework for Affect Detection-Thermal Image

Given the limited resources at our disposal, the experimental setup is illustrated in Fig.2. The experiment session was conducted in a controlled temperature and humidity room. Prior to data collection, the subjects were seated pleasantly on armchair in front of monitor as illustrated. Upon being seated, the subjects were requested to relax for 10 minutes; to allow subjects' skin surface to reach equilibrium with the temperature of the experiment room. The distance between subject and the thermal camera was maintained at 1.0 meter and the emissivity constant parameter for the thermal camera was set to 0.98 (Human skin). Throughout the session, the subjects were requested to remove their glasses if they wore any. Subjects are advised not to take any meal 1 hour before data collection to avoid metabolic effect of digestion and minimum movements, particularly above the neck were advised to the subject to avoid out of focus images. The visual stimuli were administrated through the monitor in front of the subjects. For this experiment, we used an uncooled micro bolometer Focal Plane Array (FPA) LWIR (7.5-13 µm) from FLIR (Model T420).

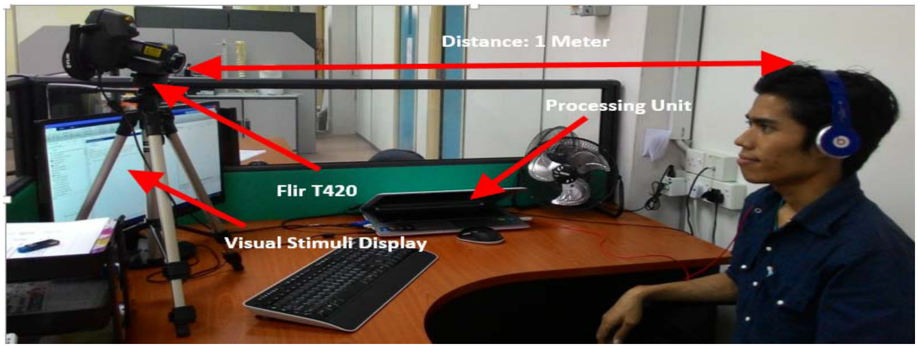

Figure 2: Setup for Data Collection

A specific movie clips have been used for the emotion induction purposes. In some literatures, International Affective Picture System (IAPS) has been proven to be effective stimuli for emotion induction. However, in our experiment movie clips been chosen over IAPS images due to the major contribution of music and sound to the emotion state as being mentioned in several literatures. In this experiment, a totally polar opposite emotion has been considered, named Happy, Sad and Fear. The duration for movie clips administered was around 1 minute each. Throughout the session, we recorded the thermal video using FLIR T420 at 30 frames per second (fps). Upon completion, the video is converted into frames.

4. Implementation of Thermal Image for Affect Detection

This section discussed on the implementation of thermal image for affect detection which involves several steps. The details are provided as follows.

4.1. Thermal Image Acquisition

The acquisition of thermal images was done in a controlled room with relative humidity of 50%, 24 °C and uniform illumination. The subject is seated 1.0 meter from the thermal camera and advice to maintain minimum head movement. The camera's framerate for data collection was set at 30 frames per second (fps). For one set of video, there are 1800 images being captured.

4.2. Image Pre-processing and Region of Interests (ROIs) Selection

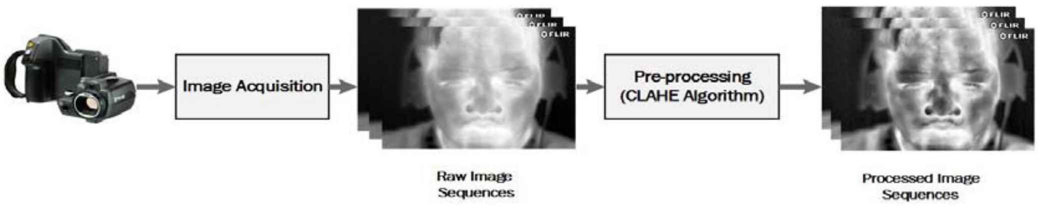

Figure 3: Image Acquisition & Pre-processing

Fig. 3 elucidates the image acquisition and pre-processing sequences. In the pre-processing stage, the raw images were normalized using contrast limited adaptive histogram equalization (CLAHE) algorithm (Bourlai, Ross, Chen & Hornak, 2012). CLAHE applies histogram equalization to each sub-region. Equation (1) explain how the histogram equalization being done. In the equation, M and N are the numbers of pixels and gray level bins in each sub-region. The histogram for each sub-region is represented as *h*. In order to improve the contrast of the image without increasing the noise, clip limit threshold is set (0.01 in our test), CLAHE redistribute each histogram so that it maintain below the threshold level. In case of gray level exceeding the clip limit and they are distributed uniformly below the threshold level. Lastly, each sub-region is combining through bilinear interpolation.

$$I(b) = \frac{(N-1)}{M} x \sum_{k=0}^{b} h(k) \tag{1}$$

Fig.4 shows the comparison before and after the implementation of CLAHE algorithm to the raw image. Based on the observation, there was significant improvement on the visibility of the hotter region on the face as represented by the white color. The algorithm enhanced the image so that the hot region could be easily identified. It utilized several distributions type to reshape the histogram of the input image, in our case we specified the distribution type to be uniformly distributed. The others two available options were 'Rayleigh' and 'Exponential' distribution.

Figure 4: Comparison between raw image and pre-processed image

4.3. Feature Extraction

In this experiment, a texture based feature extraction method was chosen (Gray Level Co-occurrence Matrix (GLCM)). GLCM was first introduced by Haralick et al. (1973) and it is one of the most popular and widely used texture feature extraction method in computer vision. The GLCM characterize second order statistic of an image by computing how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. The matrix contains the conditional joint probabilities of all pair wise combinations of gray levels given at particular displacement distance (*d*) and at particular orientation (*θ*). The displacement distance (*d*) is also known as interpixel distance. The probability (P_r) can be defined as (Barber & LeDrew, 1991):

$$\mathrm{P_r}(x) = \{C_{ij} \mid (d, \theta)\} \tag{2}$$

Where C_{ij} is defined as:

$$C_{ij} = \frac{P_{ij}}{\sum_{i,j=1}^{G} P_{ij}} \tag{3}$$

In equation (3), the variable C_{ij} represents the number of occurrences of gray level in row (*i*) and column (*j*). P_{ij} value is the probability value from the GLCM (how many times that reference value occurs in a specific combination with a neighbour pixel). *G* is the total number of gray levels. At a particular instance, at a given the displacement distance (*d*) and orientation (*θ*), the summation of the denominator in equation (3) represent the total number of gray level pairs within a window. The features are generated by calculating the features for each one of the co-occurrence matrices obtained by using the directions 0°, 45°, 90°, and 135°, then averaging these four values (Clausi, 2002). Several texture statistics can be extracted from the GLCM matrix; Contrast, Correlation, Energy, Homogeneity, etc. In our experiment, we only considered these four features from the second order statistics. In addition, for the computation of GLCM, we defined a kernel as follows which represented the displacement distance (*d*) and the orientation (*θ*).

Kernel = [0 1; 0 2; 0 3; 0 4;...
-1 1;-2 2;-3 3;-4 4;...
-1 0;-2 0;-3 0;-4 0;...
-1 -1;-2 -2;-3 -3;-4 -4;]

From the kernel, the GLCM will have four displacement distances, d= 1,2,3,4 and four orientations/angles, θ =0°, 45°, 90°, 135°. So for each GLCM statistical features, there are 16 features (16 features x 4= 64 features for single ROI). Hence, for each ROI, there are (64 features x 200 images = 12800 features). The Happy, Sad and Fear dataset have 12800 features respectively which leads to a total of 230400 features (6 subjects, all male) overall for each ROI. The dataset (EmotionTable) then represented in the form of 230400 features x 4 ROIs (Supraorbital, Periorbital, Nasal, and Mouth).

4.4. Classification

For the classification part, several Classifiers have been tested in order to achieve the best classification result. Among those were Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Tree. The size of dataset (EmotionTable) fed into the classifier was 230400x4 features. Most of the classification processes was done using Classification Learner in Matlab 2015a. The ROIs served as the predictors while the Emotion labels as the response. The 10-Fold Cross Validation was chosen for model validation.

5. Result & Discussion

From all the classifiers tested in classification stage, Weighted KNN has shown the best classification accuracy with 98.2% overall accuracy. The second highest accuracy score belong to Fine KNN with overall accuracy of 97.9% which 0.3% lowers than Weighted KNN. Besides, Fine Gaussian and Medium Gaussian SVM displays quite impressive classification result with 94% and 82.3% overall accuracy respectively. The comparisons of the classifier performance are tabulated in Table 1 below in term of overall accuracy. The detailed accuracy of the classifier is tabulated in Table 2. The AUC were obtained from the Receiver Operating Characteristic (ROC) curve. The higher the AUC value, the better the accuracy of the classifier as the ROC curve represents the sensitivity as a function of fall-out.

Table 1: Classifier Performance Comparison

Classification Method	Feature	Accuracy (%)
SVM: Fine Gaussian	GLCM	94.0
SVM: Medium Gaussian	GLCM	82.3
Tree Complex	GLCM	63.5
Tree: Medium	GLCM	49.8
KNN: Fine KNN	GLCM	97.9
KNN: Weighted KNN	GLCM	98.2

Table 2: Detailed Accuracy by Class for Weighted KNN Classifier

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area /AUC
Happy	0.975	0.013	0.974	0.975	0.975	0.997
Sad	0.972	0.012	0.975	0.972	0.974	0.997
Fear	1	0.001	0.998	1	0.999	1
Weighted Avg.	0.982	0.009	0.982	0.982	0.982	0.998

Table 2 tabulated the detailed accuracy of Weighted KNN Classifier by classes of emotions. The emotion classes are Happy, Sad and Fear. The True Positive (TP) Rate is also known as sensitivity where its measures the proportion of positives that are correctly classified as positive. The highest TP Rate from the classes of emotion belong to Fear-Class with 100% correctly classified instances. The False Positive (FP) Rate is a measures of incorrectly identified or false alarm which contribute to type I error in statistic. The smallest FP Rate could be observed from Fear-Class with 0.1%. The Precision parameter is also known as Positive Predictive Value (PPV) where its measures the proportion of the true positives against all the positive result (TR and FP). Lastly, the F-Measure is a parameter that measures the harmonic mean of precision and recall. It can be used as a single performance index for the positive class classification.

Fig.5 illustrates the confusion matrix for Weighted KNN. From the confusion matrix, several important variables could be extracted; True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), True Positive Rate (TPV), False Negative Rate (FNR), Positive Predictive Rate (PPR), False Discovery Rate (FDR) etc.

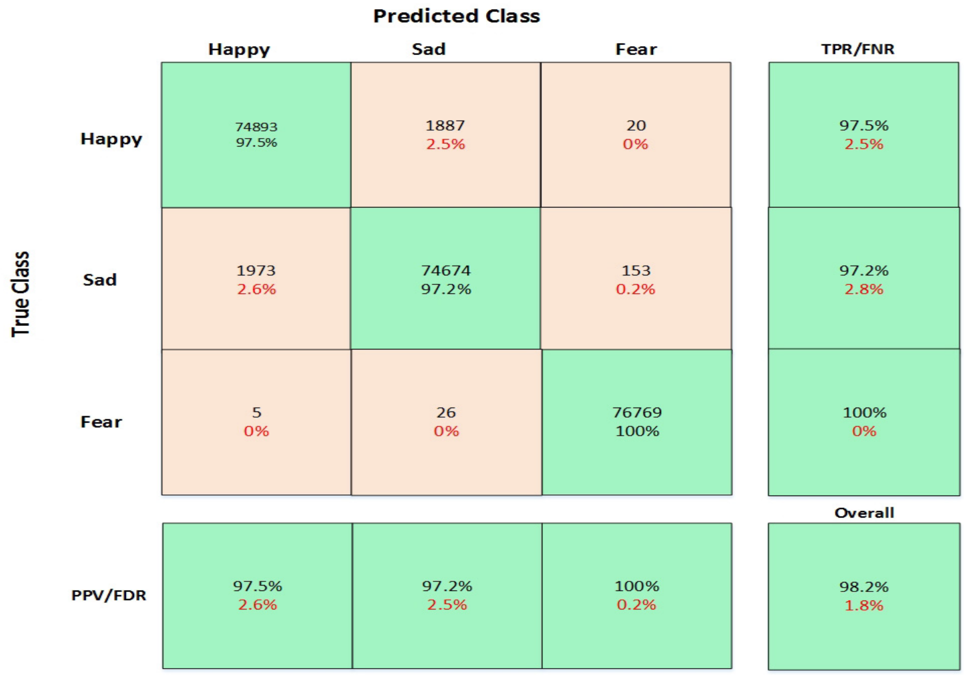

Figure 5: Confusion Matrix (Weighted KNN)

In addition to that the 10-fold cross validation performed on the Weighted KNN Classifier has resulted in 2% k-fold loss which demonstrates the small out of sample error. The small out of sample loss shows that the model performed accurately as the fitting for each fit session (10-fold) are well-fitted with minimum average error. The objective of k-fold cross validation is to estimate how accurately the predictive model would perform in practice. With the small resulted k-fold loss, we are positive and optimistic that the predictive model would perform well in real life scenario. By considering several features vector being concatenated together, the overall performance might be improved due to availability of several distinct features. Besides, the features vectors dataset could be pre-process before feed in classification phase using Principle Component Analysis (PCA) or Linear Discriminant Analysis (LDA) to reduce the dimension of the dataset. Furthermore, for real life application of thermal imaging, several known limitations exist. The environmental condition plays the crucial role since the homeostasis and the cutaneous temperature are continuously regulated. In order to avoid attributing any psychological valence to unmitigated thermoregulatory, proper countermeasures and cautions should be adopted (Merla, 2014). Besides, the occlusion of ROIs by spectacle or hair-bang could render the visibility of the thermal imprint in the targeted location. The metabolic effect of digestion also affected the regulatory of cutaneous temperature of the facial skin as reported in several studies. The model could be improved by considering the ROIs location tracker to track the ROIs which might help reducing the out of focus images. Our research is far from being conclusive. However, these findings are utmost promising. For further improvement, we would need more training data than what has been tested in this paper. This development of affective state detection from thermal imaging could be useful in HRI especially when dealing with emotionally challenged individual (e.g. Autistic, Asperger, ADHD, etc.). We believe that this affective state model could provide contactless, non-invasive emotion measurement and helps psychiatrist with emotional states evaluation and engagement with the subject.

Acknowledgements

We wish to express our gratitude to the Ministry of Higher Education (MOHE) for funding the project under the **Fundamental Research Grant Scheme (FRGS), Grant no: FRGS13-076-0317**. Our deepest appreciation also goes to postgraduate student and lecturers involved in making this project possible.

References

A. Di Giacinto, M. Brunetti, G. Sepede, A. Ferretti and A.Merla (2014). Thermal signature of fear conditioning in mild post-traumatic stress disorder. *Neuroscience* 266, 216–223, 2014

A. Merla and G.L. Romani (2007). "Thermal signature of emotional arousal: a functional infrared imaging study," in *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite Internationale*, Lyon, France, August 23-26

A. Merla (2014),"Revealing psychology and emotion through thermal imaging, "in *Proceeding of International Conference on Physiological Computing System*, 368-37

A.Merla (2013). "Advances in thermal imaging for monitoring physiology and social interaction unobtrusively, "in P*roceedings of the International Conference on Psychophysiology*,50:S14

A.Shastri, A.Merla, P.Tsiamyrtzis and I.Pavlidis (2009)." Imaging facial signs of neurophysiological response, "in *IEEE Trans. Biomed. Eng*. 56,477-484

B. Hern´andez,G. Olague, R. Hammoud, L. Trujillo,E. Romero (2007)." Visual learning of texture descriptors for facial expression recognition in thermal imagery." in *Computer Vision and Image Understanding* 106(2-3), 258–269

B. R. Nhan and T. Chau (2010). "Classifying Affective States Using Thermal Infrared Imaging of the Human Face, "in *IEEE Trans. Biomedical Eng.*, vol. 57,no. 4, 979-987

D A. Clausi (2002)."An analysis of co-occurrence texture statistics as a function of grey level quantization," Can. J. Remote Sensing, Vol. 28, No. 1, pp 45-62, 2002

D.G. Barber, E.F. LeDrew (1991). "SAR sea ice discrimination using texture statistics: a multivariate approach," Photogrammetric Engineering and Remote Sensing 57, 385–395

I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey (2007) "Interacting with human physiology," *Comput. Vis. Image Understanding*, vol. 108, no. 1–2, 150–170

I. Pavlidis, P. Tsiamyrtzi, D. Shastri, A. Wesley, Zhou, P. Lindner et al. (2012). Fast by nature–how stress patterns define human experience and performance in dexterous tasks. *Sci.Rep.* 2, 1–9

L. Zhilei and W. Shangfei (2011)."Emotion Recognition using Hidden Markov Models from Facial Temperature Sequences," Springer-Verlag Berlin Heidelberg, 240-247.

M. Coli, L. Fontanella, L.Ippoliti, and A. Merla (2007). "Multi resolution KLE of psycho-physiological signals," in *Proceedings of S.Co.2007,Book of Short Papers*, ISBN 978-886129-114-0,Venice,Padova,116–121.

M.Garbey, N.Sun, A.Merla and I.Pavlidis (2007). "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery*," IEEE Trans. Biomed. Eng*. 54, 1418-1426.

M.M. Khan, R.D. Ward and M. Ingleby (2009). "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *Trans. Appl. Percept*. 6, 1

R. Murthy and I. Pavlidis (2006)." Noncontact Measurement of breathing function," *IEEE Eng. Med. Biol.Mag*.25, 57-67.

R.M. Haralick, K. Shanmugam, I. Dinstein (1973) "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, No. 6, 610–621.

T. Bourlai, A.Ross, C.Chen and L.Hornak (2012)" A Study on Using Mid-Wave Infrared Images for Face Recognition," in Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX, *Proc. of SPIE* Vol. 8371.

T. Sheridan (1992). Tele robotics, Automation, and Human Supervisory Control. Cambridge, MA: MIT Press

V.Engert, A .Merla, J.A, Grant, D .Cardone, A. Tusche and T.Singer (2014). Exploring the use of thermal infrared imaging in human stress research, *PLos ONE*.

Yoshitomi, Y. (2010). "Facial Expression Recognition for Speaker Using Thermal Image Processing and Speech Recognition System," in *Proc. of 10th WSEAS International Conference on Applied Computer Science*, 182–186.